RETROSPECTIVE:

Memory Access Buffering in Multiprocessors

Michel Dubois

Department of Electrical Engineering - Systems
University of Southern California
Los Angeles, CA 90089-2562
dubois@paris.usc.edu

Christoph Scheurich

Intel Corporation
2200 Mission College Blvd.
Santa Clara, CA 95052-8119
christoph.e.scheurich@intel.com

At the time when the research work for the above paper was done, Michel Dubois was a starting assistant professor and Christoph Scheurich was a Ph.D. student at the University of Southern California. In this retrospective they both relate their experience on the work they did jointly more than 10 years ago.

The Advisor's Perspective

The problems of coherence and consistency have been addressed quite extensively in the past 10 years, to a point of information overload. It was not always that way. Back in 1982, the only definition of coherence was the one given by Censier and Feautrier in their landmark 1978 paper (the notion of "latest" copy) [1]. In that paper a central directory formed a bottleneck through which all cache requests had to go. It appeared to me back at that time that the definition was not sufficient in the context of a more distributed environment (even a bus-based system); moreover, as the speed gap between processor and memory widened, I realized that there would eventually be a time when memory accesses would have to be buffered in the processor to overlap them with each other and with computation.

I started thinking about these issues as I took a job in France in an R&D group in computer architecture from 1982 to 1984. During that time I dug out most of the literature that I could find on the topic. Mostly, the issues were touched lightly here and there in a few papers (clearly some people were thinking about the problems), and also there were Lamport's papers [2], totally ignored by the architecture community. If one of my colleagues from these early days in my career ever read this retrospective he will remember a presentation in which I explained in broad terms how to design a system such as the one shown in Figure 1, where each processor has a cache and an input and output memory access buffer.

Whether a multiprocessor system with caches and memory access buffers could ever work correctly in general was not at all clear, back in 1982.

When I took a position at U.S.C., I had the time to think more intently about these problems and wrote a paper for ISCA'85, which already included the definitions of strong and weak memory order. The paper was flatly rejected, which was extremely frustrating to me because I had the firm conviction that the material was important and fundamental.

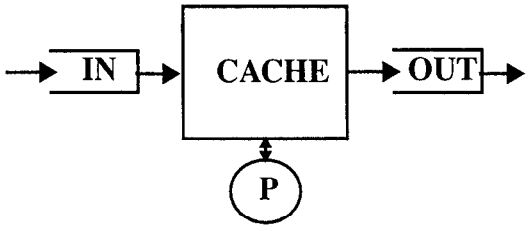

Figure 1

It appeared to me at that point that I needed help since thinking about the problem was consuming most of my time and energy and I had to teach at the same time. When I received a bit of money in the form of a research initiation grant from the National Science Foundation, I was able to start hiring students. I needed someone articulate, fearless, who has a strong sense of logic and common sense and I found that person in Christoph Scheurich who had been my teaching assistant for a semester.

After a learning period, we were able to have long technical discussions, which were very helpful. From that point progress was fast. I think we were both enthusiastic about the research. A new paper was submitted to ISCA'86. The reviews were mixed. I remember however a long, very technical review. Clearly we had connected with someone. Through the reviews we also learned of William Collier's work [3]. I remember being very relieved that someone else had been working so extensively on the topic.

Another paper was accepted the following year in ISCA'87 strictly on sequential consistency [4]. Obviously we still had problems with reviewers: I will never forget the comment of a reviewer who claimed that "he had polled his friends and no one cared about sequential consistency".

At this point more work needed to be done in refining the concepts and on performance evaluation. With another student and with the help of Faye Briggs who was then at Rice, we had developed a simple simulation environment [5]. However, the student left, and Faye Briggs went to Sun Microsystems. By that time Christoph deserved to complete his Ph.D. [6]. All my attempts at getting funding to continue the work failed. Through the vagaries of the funding process, I received some money to work on asynchronous algorithms. My focus shifted and I dropped the problem of coherence and consistency.

Meanwhile the work was still getting no recognition and again, this was extremely frustrating since I believed in the work and also there was clear interest from industry. It took until ISCA'90 before other groups published work on the same topic and finally our work started to be referenced [7][8].

The paper we published in 1986 was the first one to tie together coherence, sequential consistency and consistency relying on explicit synchronization. These issues were mostly ignored at the time, even by the large well-funded groups doing research on multiprocessors at other universities. The paper introduced the notion of general coherence (which does not rely on Censier and Feautrier's memory bottleneck) and of strong and weak memory access orders. Some of the terminology defined first in the paper such as "performed", "performed with respect to a processor", and "globally performed" has withstood the test of times and has become part of the vocabulary used to explain coherence and consistency.

The work had a large impact on both academia and industry. It ultimately led to innovations that were totally unexpected. Who could have predicted back in 1986 that architecture manuals would dedicate entire chapters to coherence and consistency and that the topic would generate the flood of ink that it did in the 1990's?

The Student's Perspective

The paper "Memory Access Buffering in Multiprocessors" was my first successful publication as a Ph.D student. However, after taking a class on data-flow computers, I was mostly interested in such architecture. It was a paper on "data-flow pipelines" that I had written for ISCA'86 that caused some unhappiness on the part of Michel Dubois (my academic advisor at USC) after he reviewed my first draft. He made some constructive suggestions but finished up by urging me to pick up on the problem of memory consistency that he had been working on before joining USC. I thought that the data-flow paper was pretty unique, I had no real understanding of what the research on memory consistency entailed, and had only mild interest in conventional multiprocessor architectures but decided that accord was in my best interest. The data-flow paper was of course summarily rejected by the ISCA referees and the research on memory consistency became the foundation of my Ph.D. thesis and also impacted other research activities outside of USC.

I read Michel's draft at least five times from the beginning to the end as well as the relevant papers by Censier and Feautrier and Lamport. When I started to understand the problem it became clear to me that thinking about parallelism involves some mental exercises that I could not perform in single steps - I needed a method to partition the problem. To that end, I spent some time with a text editor to come up with an early version of Definition 3.3 that is now part of the paper. Only later did I recognize this definition as significant. At the time I thought of it as a mental bridge.

Over a weekend I took some of Michel's earlier text and merged it with some of my own thoughts and embedded Definition 3.3. Then I went back to what I considered to be the real problem: data-flow pipelines. After Michel had some time to review my draft I expected that my stint in memory consistency research would come to a merciful end at our next scheduled meeting. I had no way of knowing that this meeting would have significant impact on my successful completion of the degree. I was thoroughly surprised when Michel's reaction to my write-up was very positive. Not only did he think that value had been added but it became obvious that he had spent some considerable amount of time thinking through the added concepts. Not surprisingly, I emerged from our first meeting on the new topic highly motivated.

In the following weeks we had many more meetings and it became clear that Michel was taking this work very seriously. We refined definitions, defined problems, and analyzed the conclusions. After we had both become fluent in the subject there were many long discussions on ordering scenarios that could become sometimes very obscure. Initially sequential consistency was our aim but then Michel shifted the focus to non-sequentially consistent models while maintaining correct operation (weak ordering). While sequential consistency made sense to us as it enforced the apparent sequential event ordering that we are all comfortable with, weak ordering initially appeared to me to be a formal definition for absolutely cha-

otic computer behavior. However, soon we managed to contain theoretically weakly ordered systems and defined rules that showed that such systems could work predictably. Because the paper is based on definitions there was very careful wording to be done as well. The simple performance models that demonstrated that weakly ordered systems can achieve higher performance than strongly ordered systems were added by Michel and made the draft ready for submission to ISCA.

The paper was accepted by the reviewers and the feedback was extremely helpful. However, one reviewer had rejected the paper outright without much elaboration. During the revision writing process we often wondered whether the negative reviewer had thought the topic mundane or whether he simply did not "get it." After some further feedback we finalized the paper hopeful that we had gotten the point across convincingly. The paper ends with the sentence: "We believe that more work is warranted in this direction." Whether or not researchers agreed with our other conclusions, that one certainly proved itself to be true. In the three years after the conference we published related research in various journals and conference proceedings [9] [10] and in 1989 my dissertation on the topic was accepted. The following year ISCA dedicated a session to the topic of multiprocessor memory models. By no means does the topic seem to be exhausted even today, as computer companies specify and build single processors and multiprocessors that vary significantly in their memory access ordering behavior.

What Happened Next?

Michel Dubois is currently a Professor in the Department of Electrical Engineering-Systems at the University of Southern California. He teaches courses on hardware design, computer architecture, parallel processing and performance evaluation. Although his mind still strays at times into the subtle intricacies of coherence and consistency his main research interests have drifted into system verification [11] and emulation of multiprocessors using FPGAs [12].

Christoph Scheurich joined Intel Corp. after receiving his Ph.D. at USC. At Intel he has worked on performance analysis, system architecture, and video capture implementations. Presently, he manages a group focused on real-time video imaging.

References

[1] L.M. Censier and P. Feautrier, "A New Solution to Coherence Problems in Multicache Systems," *IEEE Transactions on Computers*, Vol. C-27, No. 12, pp. 1112-1118, December 1978.

[2] L. Lamport, "How to Make a Multiprocessor Computer That Correctly Executes Multiprocess Programs," *IEEE Transactions on Computers*, Vol. C-28, No. 9, pp. 690-691, September 1979.

[3] W. W. Collier, *Architectures for Systems of Parallel Processes*, Technical Report TR 00.3253, IBM Corporation, Poughkeepsie, NY, January 1984.

[4] C. Scheurich and M. Dubois, "Correct Memory Operation of Cache-based Multiprocessors," *14th Int. Symposium on Comp. Arch.*, June 1987, pp. 234-243.

[5] M. Dubois, F.A. Briggs, I. Patil, and M. Balakrishnan, "Trace-driven Simulations of Parallel and Distributed Algorithms in Multiprocessors," *Proceedings of the 1986 International Conference on Parallel Processing*, August 1986, pp. 909-916.

[6] C. Scheurich, *Access Ordering and Coherence in Shared Memory Multiprocessors*, Ph.D. Thesis, Dept of EE-Systems, University of Southern California. Also Computer Engineering Technical Report No. CENG89-19, May 1989.

[7] S.V., Adve and M.D. Hill, "Weak Ordering--A New Definition," *Proc. 17th Int. Symp. on Computer Architecture*, pp. 2-14, 1990.

[8] K. Gharachorloo, et al., "Memory Consistency and Event Ordering in Scalable Shared Memory Multiprocessors," *Proc. of the 17th Int. Symp. on Computer Architecture*, pp.15-26, 1990.

[9] M. Dubois, and C. Scheurich, "Memory-Access Dependencies in Shared-memory Multiprocessors," *IEEE Transactions on Software Engineering*, Vol. 16, No. 6, June 1990, pp. 660-673.

[10] C. Scheurich and M. Dubois, "Lockup-free Caches in High-Performance Multiprocessors," *Journal of Parallel and Distributed Computing*, 11, 25-36, January 1991.

[11] F. Pong and M. Dubois, "A New Approach for the Verification of Cache Coherence Protocols," *IEEE Transactions on Parallel and Distributed Systems*, Vol. 6, No. 8, pp. 773-787, August 1995.

[12] L. Barroso, S. Iman, J. Jeong, K. Oner, K. Ramamurthy and M. Dubois, "RPM: A Rapid Prototyping Engine for Multiprocessor Systems," *IEEE Computer*, pp. 26-34, February 1995.